



Master II livello 2006

Concetti base relativi alla cella DRAM

- A) Cella di memoria DRAM costituita da un transistor e da una capacita'
 - 1) immagazzinamento dinamico di carica e refresh
 - 2) scrittura
 - 3) lettura
 - 4) obiettivi di progettazione e compromessi
 - 5) problematiche di implementazione
 - 5.1) spessore dell'isolante
 - 5.2) materiale dell'isolante
 - 5.3) incremento dell'area effettiva della capacita' (planare-trench-stack)
- B) Principi base nella operazione di sensing di una cella DRAM
 - 6) sensing del dato
 - 7) margini operativi
 - 8) elementi che contribuiscono alla definizione dei margini operativi
 - 8.1) tensione di alimentazione
 - 8.2) caduta nella tensione di soglia
 - 8.3) correnti di perdita
 - 8.4) accoppiamenti di cariche
 - 8.5) sbilanciamenti
 - 8.6) variazioni nei livelli di precarica

Naso Giovanni

in partnership with



Università degli Studi dell'Aquila



1) Immagazzinamento dinamico e refresh

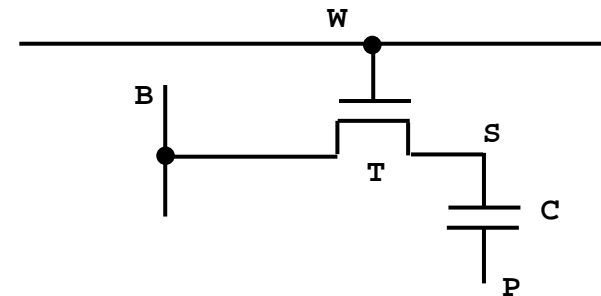
La struttura costituita da un transistor e una capacita' (1T1C) e' ormai diventata uno standard industriale per le memorie DRAM perche' :

- Occupa una piccola superficie orizzontale
- E' compatibile con un processo CMOS
- Ha buone prestazioni quando e' strutturata in un array
- Si presta facilmente ad processo di scalatura

La struttura dinamica 1T1C impiega un transistor T, una capacita' C ed il dato e' rappresentato dalla presenza o assenza di cariche immagazzinate nella capacita', ovvero dalla tensione relativa al nodo S (storage) comune tra T e C.

Quando $V_S = V_{DD}$ (tensione di alimentazione) nella cella si intende memorizzato un valore logico 1. Quando $V_S = 0$ volt, nella cella si intende memorizzato un valore logico 0.

La linea W che costituisce la gate del transistor T si chiama 'word line'; il terminale del transistor T diverso dal nodo di storage si chiama 'bit line'; il terminale P della capacita' e' il plate comune a tutte le capacita' dell'array e per semplicita' si assume essere il ground.

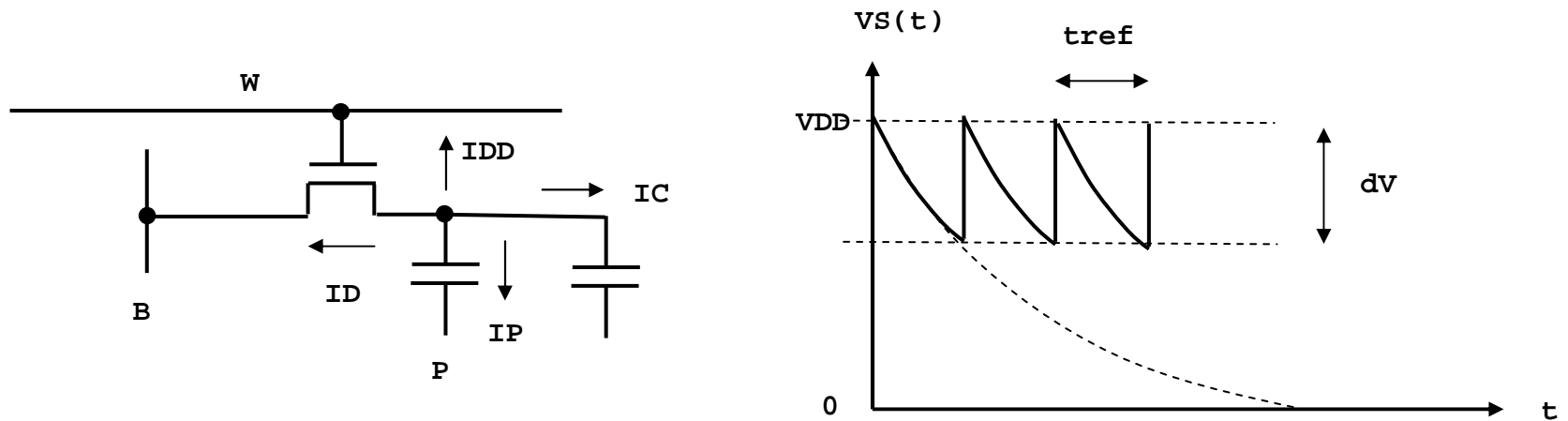


1) Immagazzinamento dinamico e refresh (cont.)

Una volta immagazzinate sul terminale S della capacita' C, con il transistor T in stato OFF, le cariche tendono a disperdersi :

- Verso l'altro piatto P della capacita' : corrente IP
- Verso le altre celle dell'array : corrente IC
- Verso l'alimentazione VDD : corrente IDD
- Attraverso il transistor di selezione : corrente ID

originando una corrente totale di leakage $IL = IP + IC + IDD + ID$.



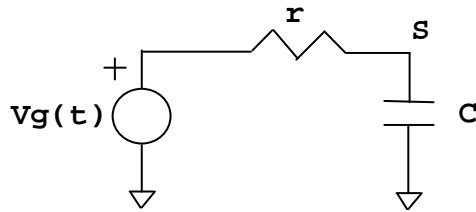
A causa della corrente di leakage, la tensione nel nodo S tende a diminuire nel tempo. Si definisce tempo di refresh t_{ref} il tempo che occorre alla corrente di leakage per variare il potenziale nel nodo di storage di una quantita' dV detta 'margine di immunita'.

$$IL = C * \frac{dV}{t_{ref}} \quad \rightarrow \quad t_{ref} = C * \frac{dV}{IL}$$

Per $C = 1fF$, $IL = 10E-14A$, $dV = 800mV$ si ha $t_{ref} = 80$ msec.



2) scrittura



La struttura di una cella DRAM 1T1C si puo' schematizzare, ai fini della operazione di scrittura, con un modello semplificato.

Se si assume che durante una operazione di scrittura :

- la word line W si porti istantaneamente ad un valore $VDD + VT$ (VT = tensione di soglia del transistore T)
- la bit line B si trovi ad un valore costante VDD
- la resistenza ON del transistore T sia costante nel tempo e uguale a r
- la capacita' di immagazzinamento costante nel tempo e uguale a C

L'andamento temporale della tensione nel node S di storage in write si puo' scrivere :

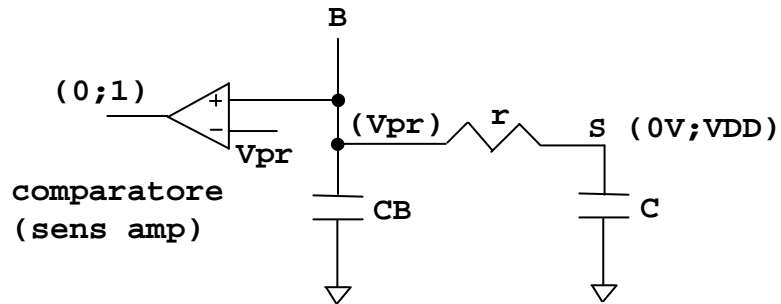
$$V_S(t) = VDD * \left(1 - e^{-\frac{t}{rC}} \right)$$

Da cui si puo' dire che :

- t_r = tempo di salita ; t_f = tempo di discesa ; $t_r = t_f = 2.3 * rC$
- t_p = ritardo di propagazione ; $t_p = 0.7 * rC$



3) lettura



Prima della chiusura del transistor :

A) 1 logico $\rightarrow V_B = V_{pr} < V_{DD} ; V_S = V_{DD}$

B) 0 logico $\rightarrow V_B = V_{pr} < V_{DD} ; V_S = 0V$

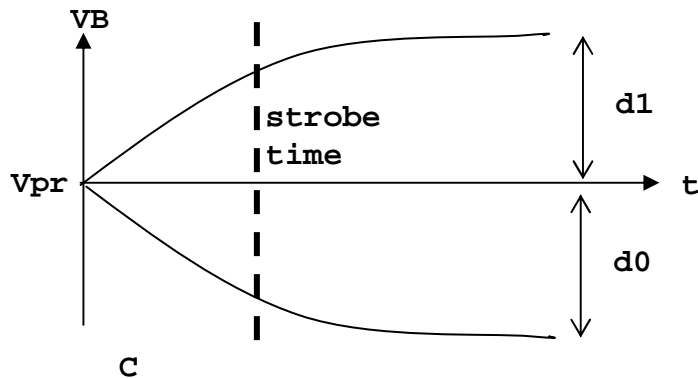
V_{pr} e' una precarica della bit line B

Dopo la chiusura del transistor (boosted gate) :

A) 1 logico $\rightarrow V_B = V_S = \frac{C \cdot V_{DD} + C_B \cdot V_{pr}}{C + C_B}$

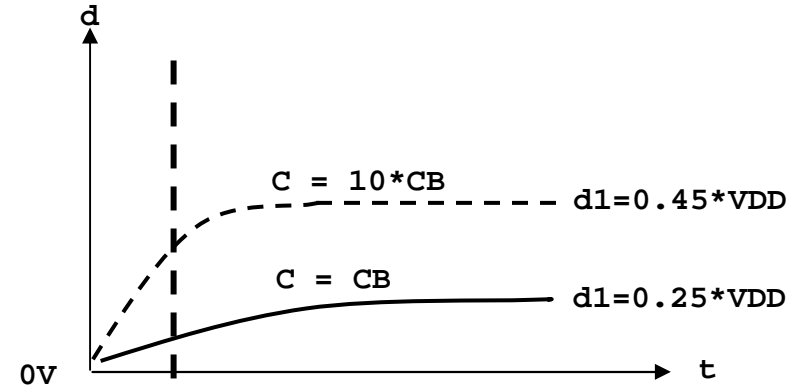
B) 0 logico $\rightarrow V_B = V_S = \frac{C_B}{C + C_B} \cdot V_{pr}$

Variazione temporale di V_B :



$$d1 = \frac{C}{C + C_B} \cdot (V_{DD} - V_{pr}) ; d0 = - \frac{C}{C + C_B} \cdot V_{pr}$$

$$|d1| = |d0| \text{ se } V_{pr} = V_{DD}/2$$



Il sense viene effettuato rilevando la variazione di V_B rispetto V_{pr} .
Se C e' molto piu' piccolo di C_B il margine di sense e' molto piccolo.

4) Obiettivi di progettazione e compromessi

Gli obiettivi di progettazione di una cella dinamica 1T1C sono molteplici ma i parametri su cui si puo' intervenire sono essenzialmente due : la capacita' C di immagazzinamento e le dimensioni W (L) del transistor di trasferimento.

C'e' inoltre la complicazione che i differenti obiettivi di progettazione richiedono valori contrastanti di C e W per cui la definizione della struttura di una cella e' frutto di compromesso :

<u>Obiettivi di progettazione</u>	<u>C</u>	<u>W</u>
Piccola area superficiale	piccola	piccola
Molte celle sulla stessa bit line	piccola	piccola
Grandi margini operativi e immunita' al rumore	grande	piccola
Lettura veloce	grande	grande
Scrittura veloce	piccola	grande
Lunghi tempi di refresh	grande	piccola
Bassi consumi di potenza	piccola	piccola
Insensibilita' all'impatto di particelle	grande	piccola
Operativita' in condizioni estreme	grande	piccola

Paradossalmente, la capacita' C di immagazzinamento in una cella DRAM 1T1C deve essere incrementata all'aumentare della densita' dei bit per area in una memoria in quanto all'aumentare della densita' aumentano accoppiamenti e leakage. Diverse vie possono essere percorse per recuperare i valori di C quando si riducono le dimensioni della capacita' di storage :

- Riduzione dello spessore dell'ossido t_d
- Uso di isolanti con alta costante dielettrica
- Diminuzione delle capacita' parassite C_{ld} che si accoppiano in serie con C
- Aumento della effettiva area AC della capacita' di storage senza aumentare la dimensione orizzontale



5.1) Problematiche di implementazione : spessore dell'isolante

Lo spessore t_d dell'isolante della capacita' di immagazzinamento in una cella DRAM 1T1C si riduce proporzionalmente con la miniaturizzazione dell'intera memoria. Questa riduzione proporzionale potrebbe non essere sufficiente ed ulteriore assottigliamento dell'isolante potrebbe essere necessario per ottenere la voluta capacita' di storage.

Una eccessiva riduzione di t_d e' associata pero' a inconvenienti di cui bisogna tener conto e che potrebbero impedire di scendere a valori troppo piccoli di t_d . Tali inconvenienti sono associati ad un eccessivo campo elettrico $E = VC / t_d$ dove VC e' la tensione ai capi della capacita'. Gli inconvenienti sono i seguenti :

- La conduzione ID attraverso l'isolante cresce vistosamente con E secondo la legge

$$ID = I_0 * e^{B\sqrt{E}} \quad \text{essendo } I_0 \text{ e } B \text{ costanti per lo specifico dielettrico}$$

- La corrente quantistica di tunnel I_{tu} attraverso il dielettrico potrebbe diventare rilevante
- La densita' di difetti D_d cresce al diminuire di t_d secondo la legge

$$D_{d2} = D_{d1} * e^{-0.2 * (t_{d2} - t_{d1})}$$



5.2) Problematiche di implementazione : materiale dell'isolante

Il materiale piu' usato come isolante nelle celle DRAM 1T1C e' il biossido di silicio SiO2.

SiO2 e' un isolante di fondamentale importanza in tutti i circuiti integrati a semiconduttore.

SiO2 e' un materiale paraelettrico; cioe' la variazione di densita' di carica dD prodotta da un campo elettrico E associato alle due tensioni V1 (livello logico 1) e V0 (livello logico 0) e' proporzionale al campo elettrico applicato e non vi e' polarizzazione spontanea P.

$$dD = \epsilon_0 * \epsilon_d * E + P \quad P = 0 ; \epsilon_d \text{ e' la costante dielettrica di SiO2}$$

Maggiore e' la costante dielettrica del materiale, minore puo' essere l'area Ac della capacita' di immagazzinamento per ottenere lo stesso effetto di storage dD.

Materiali paraelettrici con maggiore costante dielettrica del SiO2 hanno comunque parametri I0 e B maggiori, tali da produrre maggiori correnti di conduzione attraverso l'isolante.

Le caratteristiche dei tre isolanti paraelettrici che possono realisticamente essere usati nelle memorie DRAM sono riportate di seguito :

<u>Materiale</u>	<u>εd</u>	<u>I0</u>	<u>B</u>
SiO2	3.9	5.1 * 10E-28	17.7
Si3N4	7.0	9.2 * 10E-18	11.7
Ta2O5	23.0	4.1 * 10E-15	23.4



5.3) Problematiche di implementazione : incremento dell'area della capacita'

Il processo delle DRAM si e' evoluto negli ultimi anni cambiando notevolmente la configurazione della cella in modo da aumentarne il piu' possibile la superficie.

Nonostante le innumerevoli varianti e' possibile classificare le celle DRAM 1T1C in 3 grandi categorie : planare, trench e stacked (riportate nel foglio successivo).

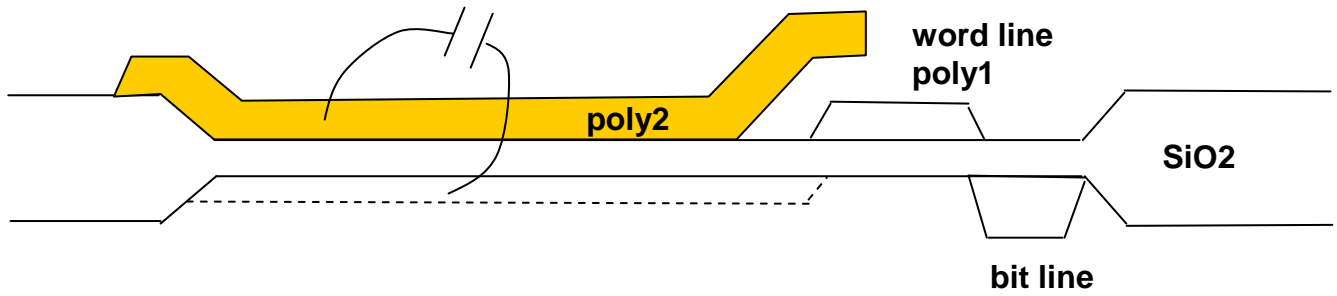
Nella cella planare, la bit line corre perpendicolarmente al disegno ed e' composta da una linea collegante tutte le drain dei transistori di colonna. La word line, realizzata da un primo livello di poly (poly1) corre parallelamente al disegno. Una delle armature del condensatore e' formata da un segmento di poly di secondo livello (poly2) normalmente connesso ad una tensione positiva in modo da invertire il substrato sottostante che costituisce sia la source virtuale del transistore di selezione, che l'altra armatura del condensatore.

Nella cella trench il condensatore e' ottenuto scavando con attacco chimico anisotropico il substrato e ricoprendo le pareti interne di SiO₂. E' possibile controllare accuratamente la profondita' (alcuni micron) e la pendenza (alcuni gradi) delle pareti del trench, oltre che la qualita' e lo spessore uniforme del dielettrico. Il poly2 connesso alla tensione positiva si insinua nel trench mentre la source virtuale circonda le pareti del trench.

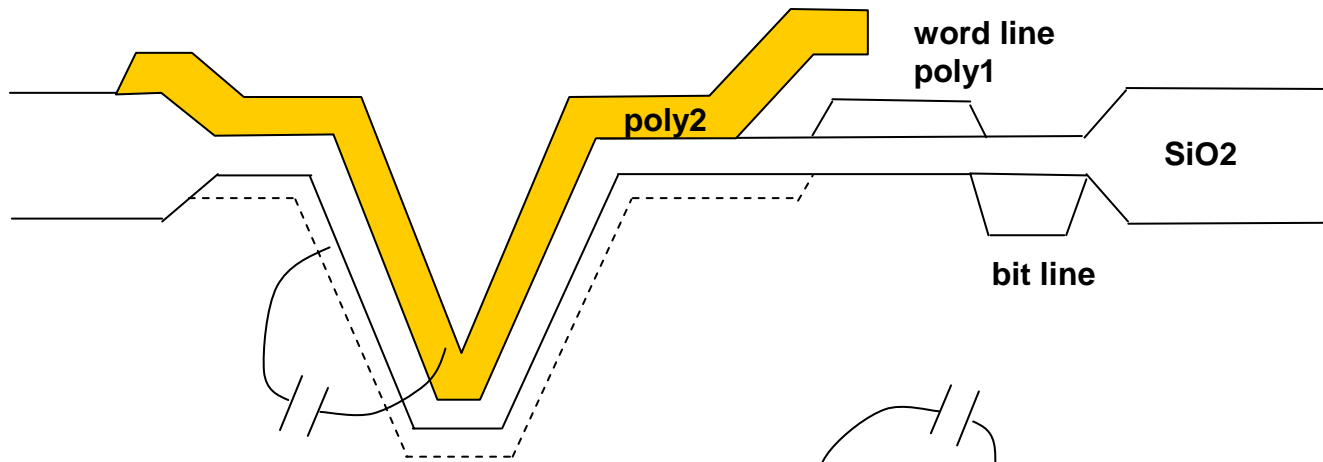
Nella cella stack la capacita' e' realizzata introducendo un terzo livello di poly (poly3). In questa struttura e' piu' semplice inserire tra poly2 e poly3 un dielettrico con alta costante dielettrica diverso da SiO₂, aumentando la capacita' a parita' di area.



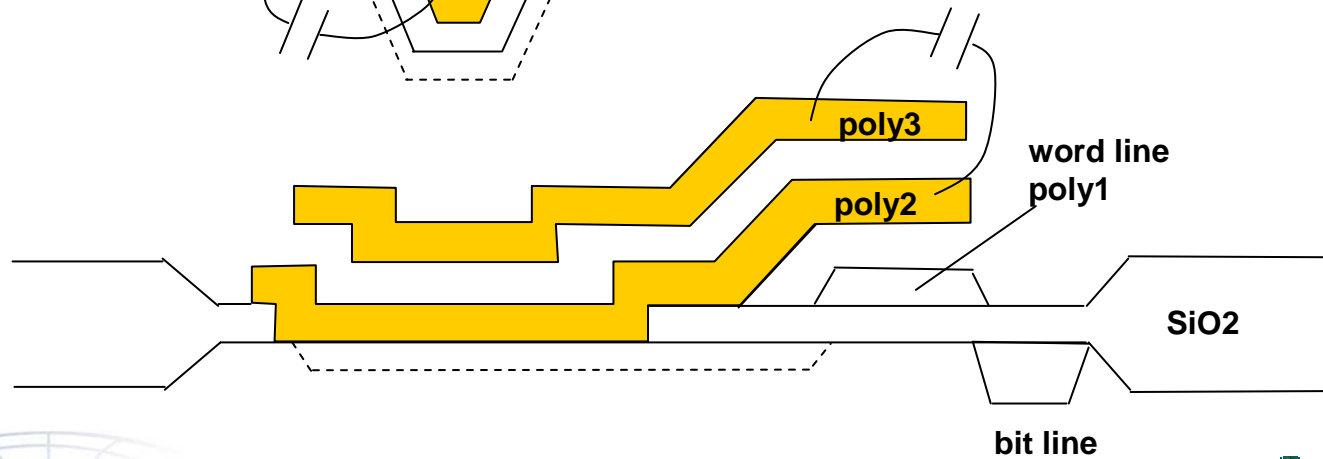
5.3) Problematiche di implementazione : incremento della'area della capacita' (cont)



Cella DRAM planare



Cella DRAM trench



Cella DRAM stack

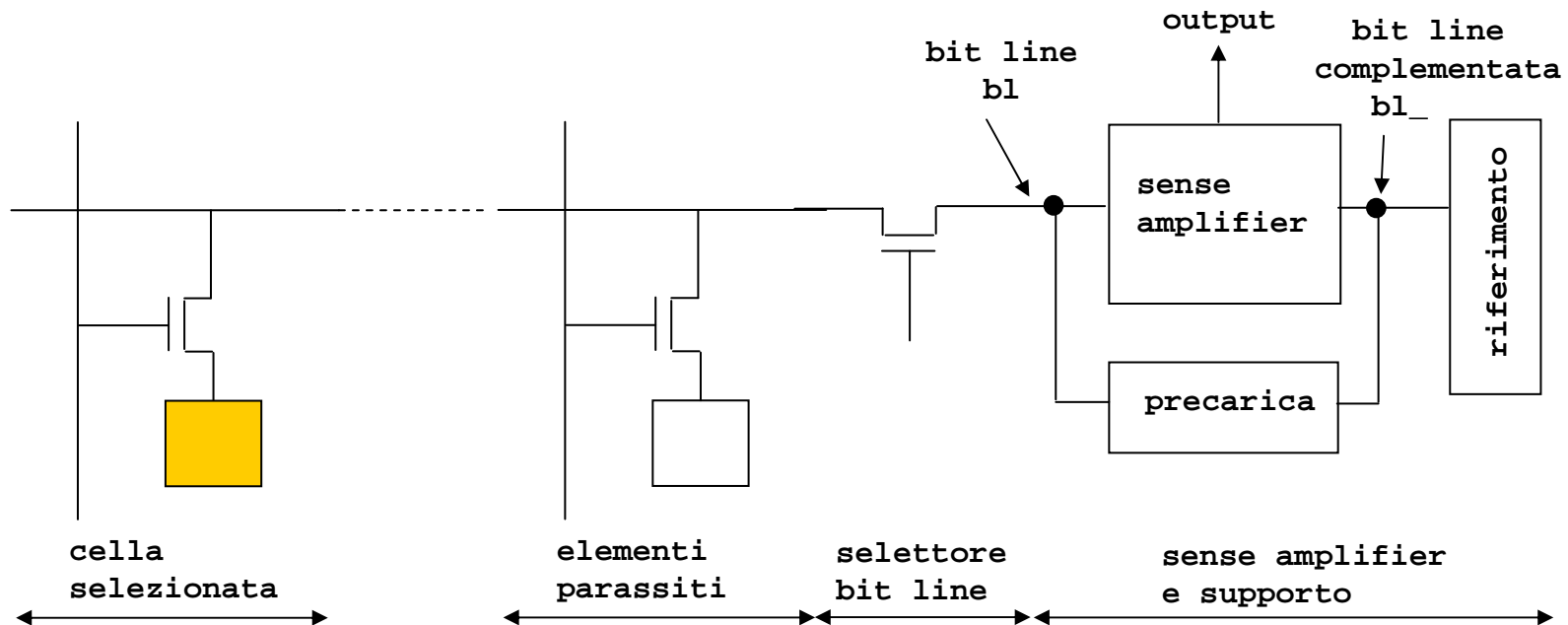


6) Sensing del dato

In una memoria integrata, 'sensing' significa rilevamento e determinazione del dato contenuto in una cella selezionata. Il sensing puo' essere non-distruttivo quando il dato contenuto in una cella non puo' essere alterato da una operazione di sensing (SRAM, ROM, EPROM) oppure distruttivo o rigenerativo quando il dato contenuto nella cella selezionata puo' essere alterato dalla operazione di sensing (DRAM).

Il sensing e' effettuato con strutture circuitali e/o parassite che tipicamente sono costituite da :

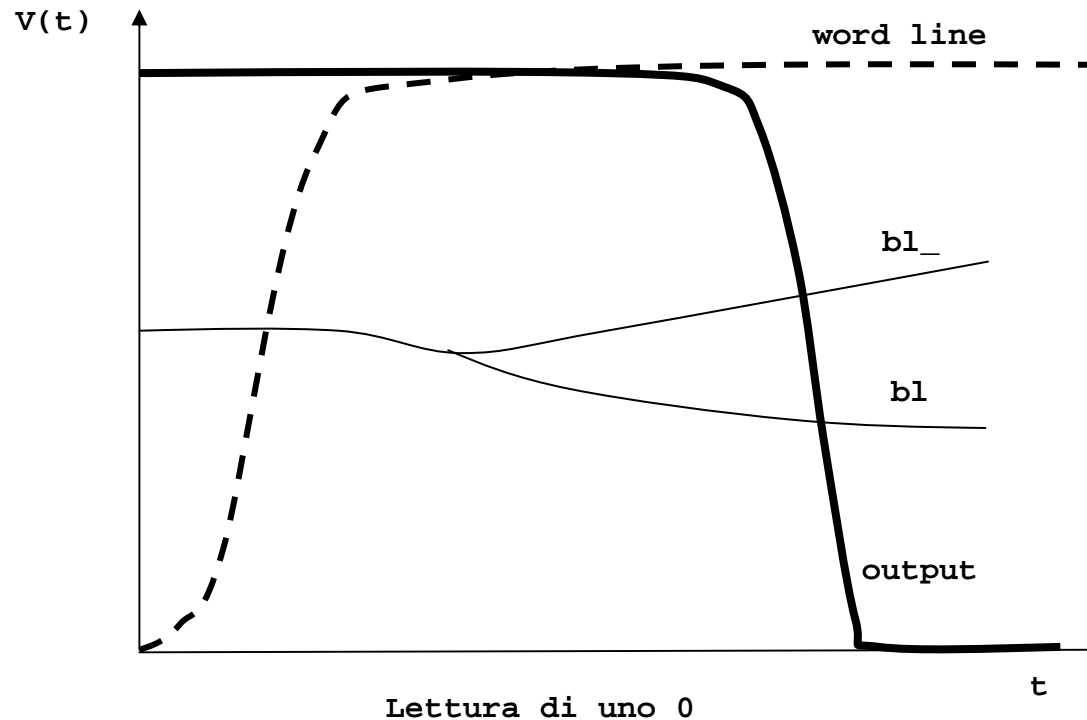
- sense amplifier
- circuiti di supporto al sense amplifier (precarica, riferimenti e circuiti di carico)
- selettori di bit line
- la cella che deve essere letta
- elementi parassiti (resistenza e capacita' della bit line e carico delle celle non selezionate ma connesse alla bit line)



6) Sensing del dato (cont.)

A causa del fatto che la resistenza e la capacita' della bit line sono grandi e l'energia di uscita di una cella di memoria e' piccola, una cella selezionata puo' generare solamente piccole correnti e piccole variazioni di tensione.

Questi segnali hanno tempi di propagazione lunghi e ampiezza insufficienti a generare il minimo valore per uno 0 logico o il massimo valore per un 1 logico. Per migliorare la velocita' di lettura di una memoria e per generare segnali di ampiezza compatibile con i circuito di periferia di una memoria si usano i sense amplifiers.



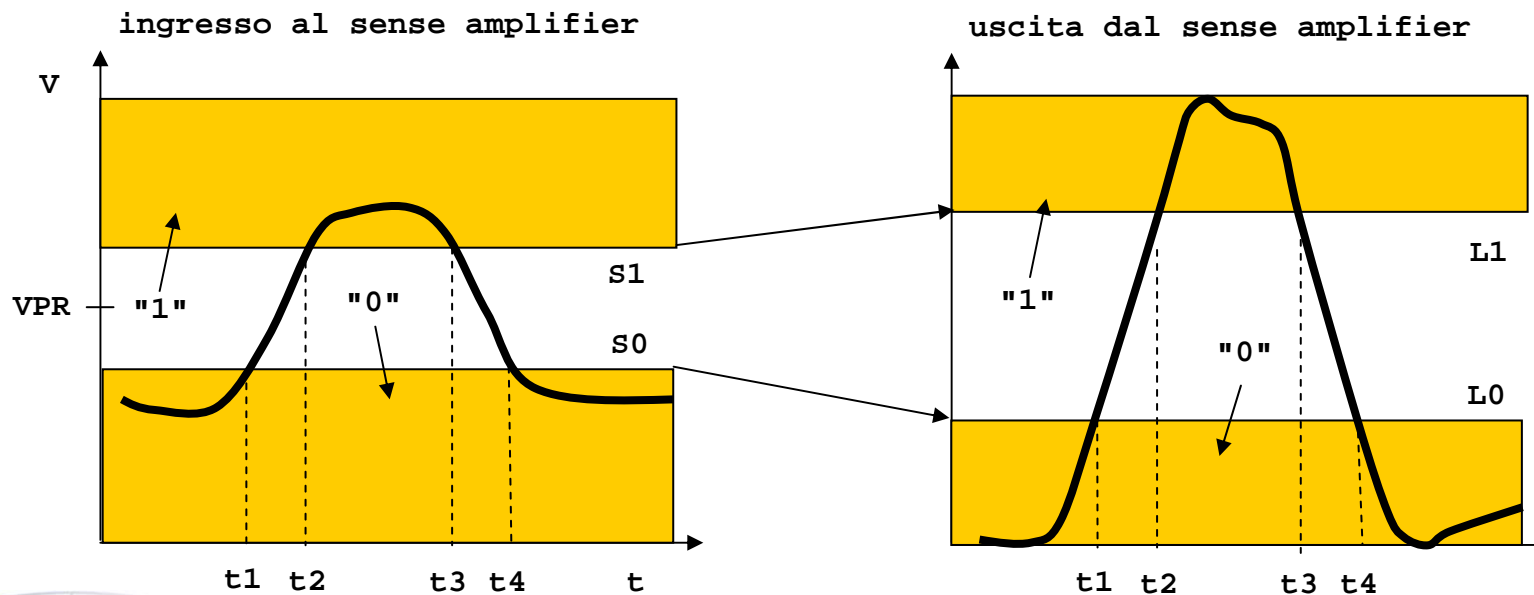
7) Margini operativi

I margini operativi in un circuito digitale sono quelle fasce di tensioni e correnti che definiscono senza ambiguita' i dati per tutte le condizioni operative del circuito.

la definizione dei margini operativi e' influenzata da :

- architettura del circuito (configurazioni circuitali, dimensione componenti, carichi ingresso/uscita, elementi parassiti)
- variazioni di processo
- condizioni di funzionamento (tensioni di alimentazione, temperatura, radiazioni)

In generale i margini operativi all'ingresso di un sense amplifier sono differenti da quelli richiesti per pilotare i circuiti logici di periferia. Un sense amplifier deve amplificare e traslare i piccoli margini operativi al suo ingresso in margini piu' ampi richiesti per il funzionamento dei circuiti logici periferici.



7) Margini operativi (cont.)

La relazione tra i margini operativi del sense amplifier e della logica periferica impongono che :

- le tensioni sulle bit lines devono poter essere minori di S_0 e maggiori di S_1
- se si vuole una simmetria tra tempi di rilevazione di 0 e 1, il valore VPR di precarica deve essere $VPR = (S_1 + S_0)/2$
- L'amplificazione A del sense amplifier deve essere $A = (L_1 - L_0) / (S_1 - S_0)$

Nella progettazione di un sense amplifier, i margini operativi sono influenzati da fattori elettrici :

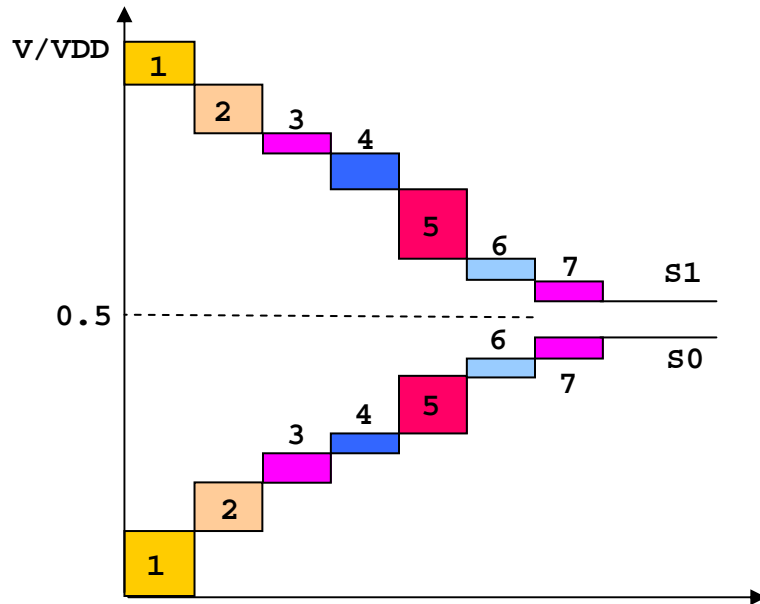
- 1) tensione di alimentazione
- 2) cadute delle tensioni di soglia
- 3) correnti di leakage
- 4) accoppiamenti di cariche
- 5) sbilanciamenti
- 6) altri effetti specifici
- 7) variazioni nei livelli di precarica

Questi fattori elettrici a loro volta hanno variazioni dovute a fattori fisici :

- a) processo
- b) temperatura
- c) condizioni di polarizzazione
- d) radiazioni

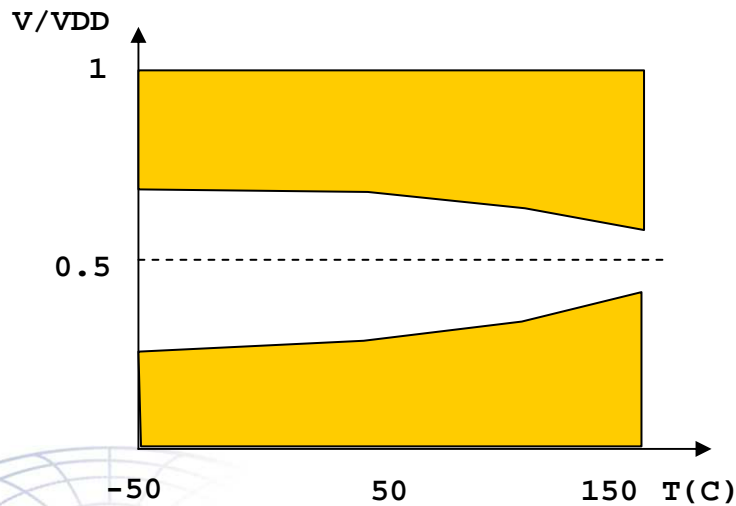


7) Margini operativi (cont.)



A causa delle variazioni fisiche a,b,c,d, ognuno Dei fattori elettrici 1, ...,7 ha un range di variazione che contribuisce ad erodere i margini di sensing.

La valutazione di possibili valori estremi per tutte le grandezze potrebbe portare alla sovrapposizione delle fasce di 1 e 0 con l'annullamento dei margini. E' opportuno quindi definire limiti per ognuna delle condizioni riportate in modo da garantire opportuna sicurezza nei margini di sensing.



A parita' di condizioni elettriche 1,...7, i margini di sensing variano per ognuna delle condizioni fisiche a,...,d.

Esempio tipico e' il margine operativo della tensione di alimentazione VDD in funzione della temperatura T.



8) Elementi che contribuiscono alla definizione dei margini operativi

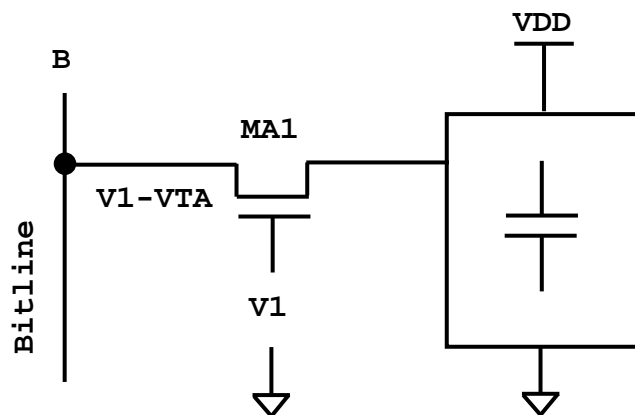
La valutazione dei principali fattori che determinano i margini operativi di un circuito di sensing deve basarsi sui valori dei parametri elettrici che peggiorano la situazione (worst case), cioè che riducono al massimo il margine di rilevamento di "0" oppure "1" di un circuito di sensing.

8.1) tensione di alimentazione

La tensione di alimentazione influisce notevolmente sui margini operativi. Il caso peggiore si ottiene in corrispondenza alla minima tensione di alimentazione dichiarata nelle specifiche (per esempio VDD-5%).

8.2) la caduta della tensione di soglia

La caduta massima di tensione VTA attraverso il transistor di accesso MA1 di una cella DRAM riduce i margini di sensing dello "0" (MA1 pch) o dell' "1" (MA1 nch).



Se per semplicità si assume il transistor di accesso di tipo nch, la massima tensione di sensing ottenibile sulla bit line varia da $V_B=V_{DD}$ a $V_B=V_1-V_{TA}$ dove V_1 è la tensione associata al livello logico "1" nel circuito e V_{TA} è la massima caduta di tensione attraverso MA1.

V_{TA} è funzione della polarizzazione del substrato VBG, della temperatura T e della dose di radioattività D .



La variazione di soglia VTA puo' essere espressa in prima approssimazione lineare come:

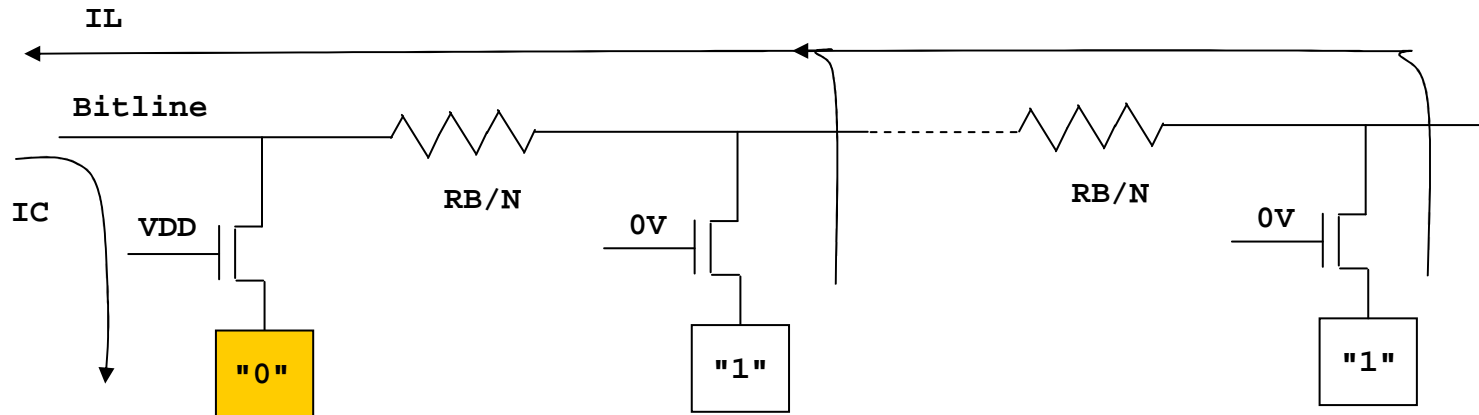
$$VTA(VT0, VBG, T, D) = VT0 + DVT(VBG) + DVT(T) + DVT(D) \text{ essendo :}$$

- $VT0$ la massima tensione di soglia con $VBG=0$, $T=25C$
- $1/2$
- $DVT(VBG) = K * (2F + VBG)$ K e' una costante e F la funzione di Fermi
- $DVT(T) = Kt/T$ Kt e' il coefficiente lineare di temperatura $2.4mV/C$
- $DVT(D)$ e' fortemente non lineare e il suo valore puo' essere abbastanza grande da azzerare i margini di sensing

Se $VTA(VT0, VBG, T, D)$ dovesse essere eccessivamente elevata, si puo' cercare di annullarla o ridurla notevolmente innalzando il valore locale di $V1$.

8.3) correnti di perdita (leakage)

La corrente di leakage IL riduce sia il margine di sensing di uno "0" che di un "1", dal momento che riduce l'ampiezza del segnale sulla bitline della quantita' $IL * RB$ e degrada il livello di tensione immagazzinata nella cella della quantita' $IL * RD$ (RB e' la resistenza lungo la bitline e RD e' la resistenza attraverso la cella).



Il caso peggiore di degradazione per leakage si manifesta quando delle N celle associate ad una stessa bitline, una memorizza un valore e le altre N-1 memorizzano un valore opposto. Per esempio : la corrente IC associata alla cella selezionata, e contenente uno "0", e' aumentata dalle N-1 componenti di corrente associate alle celle deselezionate e contenenti un "1".

Le principali correnti di leakage sono : la corrente di sub-threshold IST, la corrente di leakage di giunzione IJ e le eventuali correnti di leakage dovute a radiazioni IR.

$$IL = (IST + IJ + IR) * (N - 1)$$

La corrente di sub-threshold IST e' una funzione che a parita' di concentrazione di drogaggio e spessore dell'ossido dipende in modo esponenziale da $(VGS - VT)/T$.

La corrente di leakage di giunzione IJ e' funzione dell'area e della profondita' della giunzione e dipende dalla temperatura secondo la legge

$$T^{3/2} * e^{-1.21/KT}$$

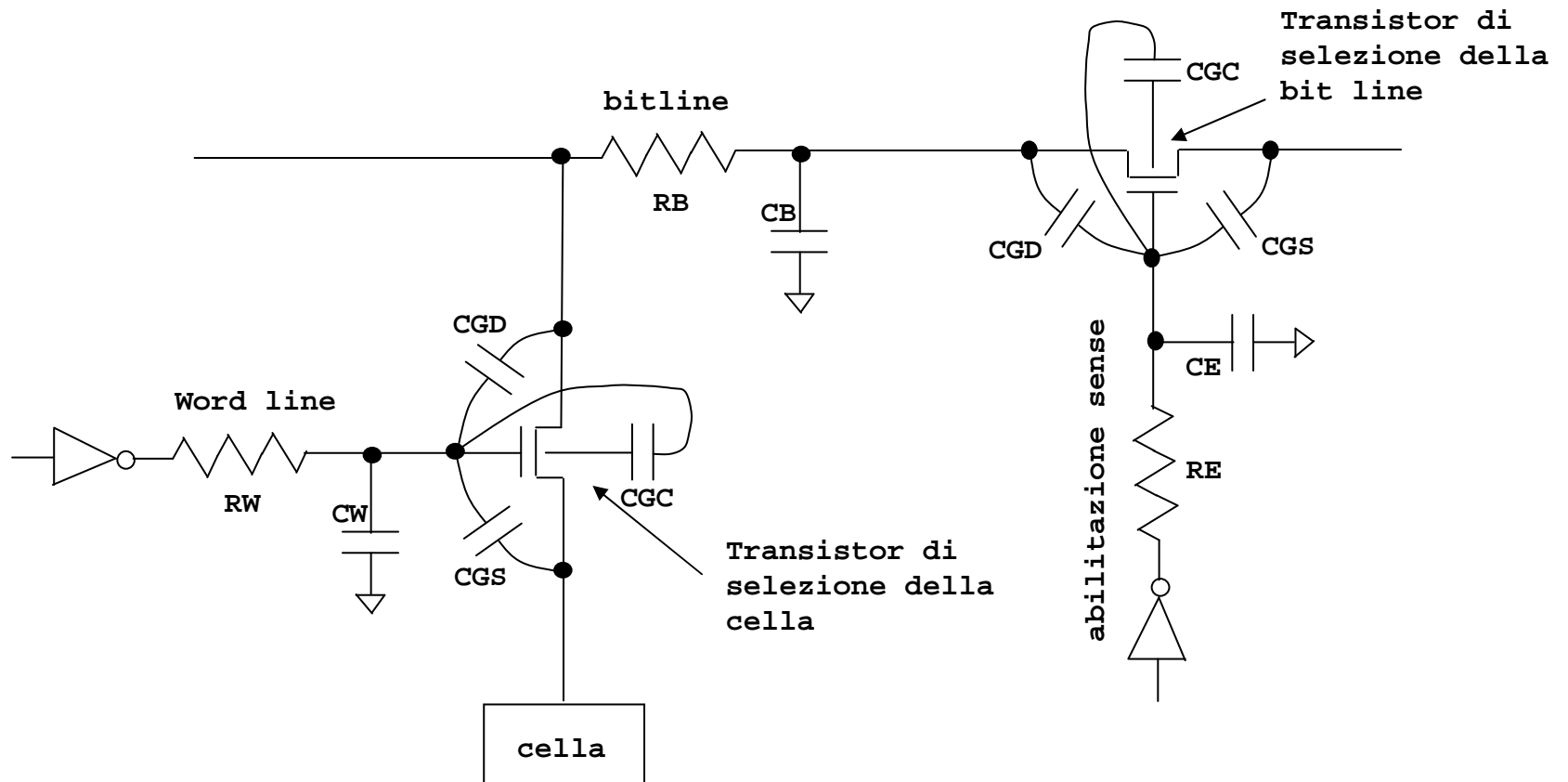
8.4) accoppiamenti di cariche

Gli accoppiamenti di cariche avvengono attraverso le capacita' CGS (gate-source), CGD (gate-drain), CGC (gate-channel) del transistor di selezione della cella DRAM e del transistor di connessione della bit line al sense amplifier.

Gli accoppiamenti di cariche influenzano sia il margine di sensing dello "0" che dell' "1" e dipendono anche dalla resistenza e capacita' della word line (RW,CW), della bitline (RB,CB) e del circuito di selezione del sense amplifier (RE,CE).

Una precisa valutazione di questi effetti richiede run di simulazioni SPICE in varie condizioni di processo, temperatura e alimentazione.



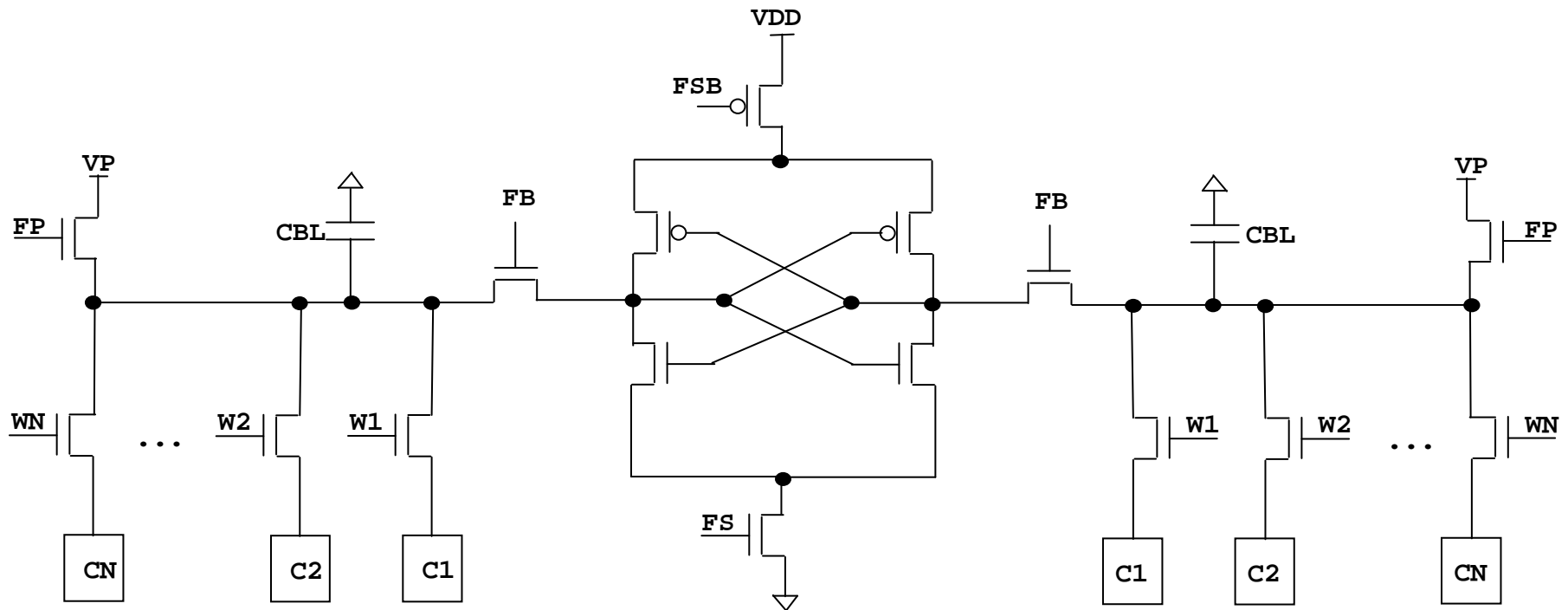


8.5) sbilanciamenti

Gli sbilanciamenti che causano degradazione nei margini di sensing sono tipici delle architetture in cui vengono usati amplificatori differenziali e degradano sia il sensing di "1" che il sensing di "0".

Il termine sbilanciamento si riferisce ad una non uniforme distribuzione topologica dei parametri nei transistors e nelle interconnessioni che costituiscono un circuito differenziale di sensing.





Idealmente un circuito di sensing differenziale e' progettato per essere topologicamente ed elettricamente simmetrico, nel senso che le due meta' del sense amplifier, la coppia di bitlines, i carichi di bitlines, i parassiti sono immagini speculari l'uno dell'altro.

Malgrado gli sforzi per avere le due meta' esattamente speculari, gli effetti delle variazioni di processo, le tensioni di polarizzazione, le correnti di substrato, la temperatura e le radiazioni possono causare una piccola e non uniforme variazione della tensione di soglia V_T , del guadagno, della capacita' di bitline e delle capacita' di giunzione e di canale da produrre come effetto totale un fenomeno di offset nel processo di amplificazione.

8.6) variazioni nei livelli di precarica

Le variazioni di livello di precarica (rispetto a quanto fissato in sede di progettazione) dovute a variazioni di processo o a variazioni ambientali producono degradazioni nei margini di sensing di uno o entrambi i livelli di "0" o "1".

Il desiderato livello di precarica puo' essere ottenuto a partire dalla tensione di alimentazione usando un circuito di partizione resistivo o capacitivo oppure un circuito che usa transistori MOS.

In un circuito di partizione, si possono avere variazioni nella tensione di precarica dovute a variazioni delle impedenze :

$$dVPR = \frac{Z1*dZ2 - Z2*dZ1}{(Z1 + Z2)^2} * VDD$$

In un circuito facente uso di MOS, la variazione di tensione di riferimento puo' essere dovuta a variazioni nella tensione di soglia del MOS e della variazione della impedenza di carico Z2.

